Mathematical Derivation of Loss Function for Logistic Regression

For EIE417 of SASTRA University (2024-25), compiled by
Suneesh Jacob

June 26, 2025

Logistic Regression Model

- Probability of class 1:

$$P(y = 1 \mid \mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + b)$$

- Sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

- Probability of class 0:

$$P(y = 0 \mid \mathbf{x}) = 1 - \sigma(\mathbf{w}^T \mathbf{x} + b)$$

Likelihood Function

- Given training set $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$
- Likelihood of the data:

$$L(\mathbf{w}, b) = \prod_{i=1}^{N} P(y_i \mid \mathbf{x}_i)$$

- Expanded form:

$$L(\mathbf{w}, b) = \prod_{i=1}^{N} \sigma(\mathbf{w}^T \mathbf{x}_i + b)^{y_i} \left(1 - \sigma(\mathbf{w}^T \mathbf{x}_i + b)\right)^{1-y_i}$$

Log-Likelihood Function

- Logarithm of the likelihood function:

$$\ell(\mathbf{w}, b) = \log L(\mathbf{w}, b)$$

- Expanded form:

$$\ell(\mathbf{w}, b) = \sum_{i=1}^{N} \left[y_i \log \sigma(\mathbf{w}^T \mathbf{x}_i + b) + (1 - y_i) \log(1 - \sigma(\mathbf{w}^T \mathbf{x}_i + b)) \right]$$

Loss Function

- Negative log-likelihood:

$$J(\mathbf{w}, b) = -\ell(\mathbf{w}, b)$$

- Loss function:

$$J(\mathbf{w}, b) = -\sum_{i=1}^{N} \left[y_i \log \sigma(\mathbf{w}^T\mathbf{x}_i + b) + (1 - y_i) \log(1 - \sigma(\mathbf{w}^T\mathbf{x}_i + b))\right]$$

Simplification Using Sigmoid Function

- For the sigmoid function $\sigma(z)$:

$$\log \sigma(z) = -\log(1 + e^{-z})$$

- and

$$\log(1 - \sigma(z)) = -\log(1 + e^{z})$$

- Therefore, the loss function becomes:

$$J(\mathbf{w}, b) = \sum_{i=1}^{N} \left[y_i \log(1 + e^{-\left(\mathbf{w}^T \mathbf{x}_i + b\right)}) + (1 - y_i) \log(1 + e^{\mathbf{w}^T \mathbf{x}_i + b}) \right]$$

Conclusion

- The logistic loss function is convex.
- It can be minimized using gradient descent or other optimization techniques.